

MACHINE TRANSLATION SYSTEM

Publication number: JP5040782

Publication date: 1993-02-19

Inventor: NARITA MASUMI

Applicant: RICOH KK

Classification:

- international: G06F17/27; G06F17/28; G06F17/27; G06F17/28;
(IPC1-7): G06F15/38

- European:

Application number: JP19910217947 19910802

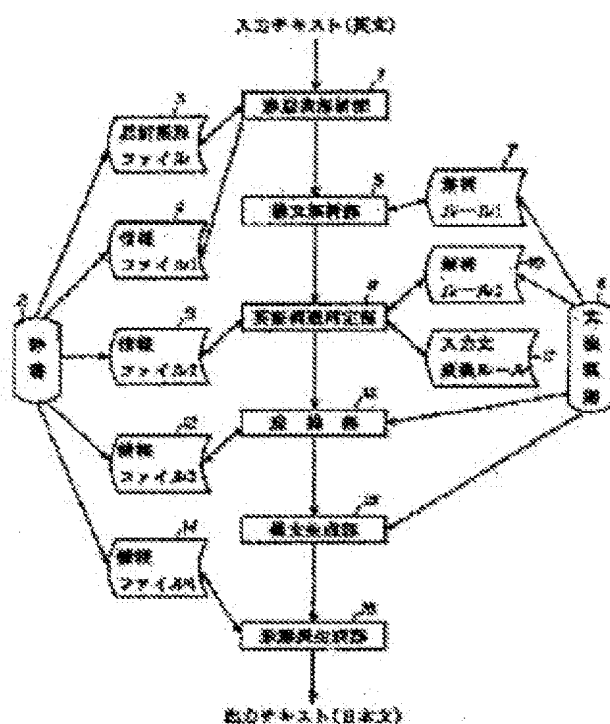
Priority number(s): JP19910217947 19910802

Report a data error here

Abstract of JP5040782

PURPOSE: To generate a more natural translated sentence to improve the easiness of understanding the translated sentence.

CONSTITUTION: An inputted English sentence is analyzed with one word as the unit by a morpheme analyzing part 1. Grammatical rules (first analysis rule 7) are successively applied to take out all allowable interpretations for the purpose of detecting phrase constitution of words, modification relations among phrases and phrases or words, or the like by a syntax analyzing part 5. Next, semantically unnatural interpretations are excluded from these syntactical analysis results in an English structure identifying part 8 by a second analysis rule 10. In a part of the English structure identifying part 8, an input sentence conversion rule 11 is accessed with the main verb in the input English sentence and its fine classification as the key to convert the input sentence if the semantic origin of the nominative case in the input English sentence is [-ANIMATE], namely, an inanimate object.



Data supplied from the esp@cenet database - Worldwide

(19) 日本国特許庁 (J P)

(32) 公開特許公報 (A)

(11) 特許出願公開番号

特開平5-40782

(43) 公開日 平成5年(1993)2月19日

(51) Int.Cl.⁵

G 0 6 F 15/28

識別記号

庁内整理番号

F 1

技術表示箇所

M 9194-5L

C 9194-5L

審査請求 未請求 請求項の数2(全7頁)

(21) 出願番号 特願平3-217947

(22) 出願日 平成3年(1991)8月2日

(71) 出願人 000006747

株式会社リコー

東京都大田区中馬込1丁目3番6号

(72) 発明者 成田 真澄

東京都大田区中馬込1丁目3番6号 株式会社リコー内

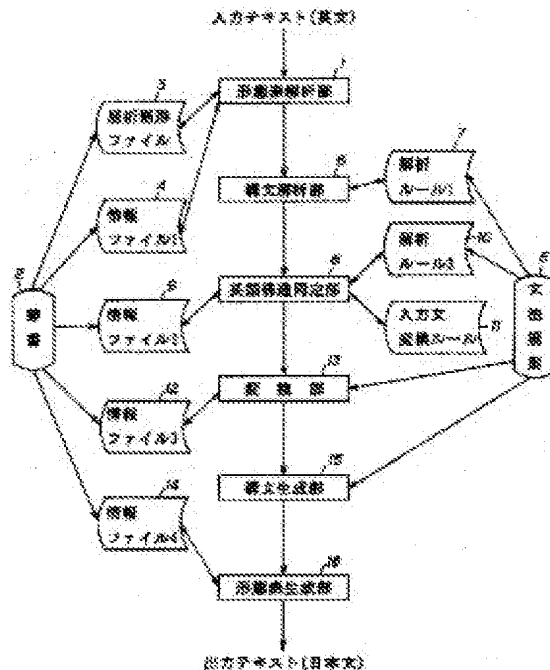
(74) 代理人 弁理士 高野 明彦 (外1名)

(54) 【発明の名称】 機械翻訳システム

(57) 【要約】

【目的】 より自然な訳文を生成し、訳文の理解容易性を向上させる。

【構成】 入力された英文は、まず形態素解析部1で単語単位の解析が行われる。次に、構文解析部5によりどの単語とどの単語で句をなし、どの句がどの句、あるいは単語を修飾しているか等を知るために、文法規則(第1の解析ルール7)を順次適用して全ての可能な解析が取り出される。続いて、このような純語論的解析結果から、英語構造同定部8により第2の解析ルール10を用いて意味的におかしい解析が排除される。英語構造同定部8中の一部として、入力英文の主語が持つ意味素性が「-ANIMATE」つまり無生物であれば、当該文中の主動詞とその經分類をキーとして入力文変換ルール11にアクセスして入力文変換を行う。



1

【特許請求の範囲】

【請求項1】 対象言語の原文を入力するための入力部と、該入力部から入力された翻訳対象となる原文を記憶する原文記憶部と、翻訳処理に使用する知識情報を収容した翻訳辞書部と、該翻訳辞書部の内容を用いて前記入力された原文の翻訳処理を実行する翻訳部と、該翻訳部の翻訳処理の結果得られた訳文を記憶する訳文記憶部と、前記原文および訳文を表示する表示部と、これらを制御するとともに、前記表示部で表示された原文または訳文の編集処理を行う編集制御部とを備えた機械翻訳システムにおいて、前記翻訳部に自然な訳文を生成するために、原文である入力文の解析結果から得られる統語的情報と意味的情報に加えて変換ルールを用いて入力文を変換する変換手段を設け、該変換手段で変換された入力文を翻訳するようにしたことを特徴とする機械翻訳システム。

【請求項2】 前記変換ルールは、主語の意味素性から無生物主語に関するものを扱うようにしたことを特徴とする請求項1記載の機械翻訳システム。

【発明の詳細な説明】

【0001】

【技術分野】本発明は、機械翻訳システムに関し、より詳細には、入力文（原文）の変換処理に係る機械翻訳システムに関する。

【0002】

【従来技術】従来の機械翻訳システムにおいては、人間の手を介さずに適切な訳文を確実に得ることは困難であり、通常、翻訳の結果として得られた複数の訳文候補をユーザに提示してユーザに最も適切な訳文を選択してもらい、その訳文候補を出力するという手続きをとっている。それでも、機械翻訳システムの訳文は直訳調であることが多く、とても自然な訳文とはいえない。特に、入力言語と出力言語で発想法が異なる場合には、直訳調の訳文では奇異に感じられ、理解容易性も大幅に低下することが多い。例えば、英語と日本語という言語対を考えてみると、英語には無生物主語の文が頻繁に使用されるが、日本語ではあまり使用されない。従来の英日機械翻訳システムでは、こういった言語間の相違を考慮せずに無生物主語の英文をそのまま日本語に置き換えてしまうので、出来上がった日本語文は不自然なものになる。例えば、"Five minutes' walk will take me to the station."なる英文が入力されると、「5分歩行は駅へ私を連れていく」というような奇異な訳文が出力される。この場合、「5分歩けば私は駅に着く」というような訳文であれば日本語として自然な文だといえる。

【0003】より自然な訳文を生成するという点に関し、これまでいくつか提案されている。例えば、①特開昭64-21262号公報の「文書処理装置」は、英語文書の作成において、作成した文書中の能動文を受動文に変形したり、受動文を能動文に変形したりするため

2

に、文型変形の対象文の構文解析を行う手段と、この構文解析の結果に基づき、オペレータから指定された文型への構文変換を対象文に施すことにより文型変形文を作成する手段とを備え、英語の受動態を日本語では能動態で訳出するといった指定を翻訳処理に入る前に行うものである。

【0004】また、②特開昭63-211464号公報の「翻訳処理方式」は、原文を翻訳する際に、文の主語や目的語にあたる部分が文構成をなして全文に比べて長すぎる原文を翻訳する際、この原文に含まれる用言に対して、形式的な主語を有する特殊な構文処理を施すための構文情報を付与する事によって原文の解析結果を前記の特殊な構文処理を施すために適した形に変換し、この結果をもとに訳出するものである。また、形式的な主語を有する構文を主語が長い構文に変換する処理を行うものとして、例えば、「英日機械翻訳システムSHALTにおける英日トランスファー」（自然言語処理研究会53-4, 1986）がある。

【0005】前記公報①の例は、日本語と比べて英語では受動態の文が多いという特徴に着目しているが、主語の意味素性を利用することは考えておらず、十分とはいえない。また、前記公報②の例に関しても、訳文の理解容易性を考慮するというよりもむしろ、日英間の文体的な相違のみに着目した変換であり、意味素性を利用した変換には到っていない。従って、いずれの場合も上記の例文のように、言語間の発想上の相違を色濃く反映したものはうまく取り扱えないという欠点がある。

【0006】

【目的】本発明は、上述のごとき実情に鑑みてなされたもので、入力文の解析処理が終了した段階で、解析結果が持つ統語的かつ意味的情報に基づいて、変換規則を適用して入力文（原文）を変換し、この変換された入力文を翻訳することにより、より自然な訳文を生成することができるようにした機械翻訳システムを提供することを目的としてなされたものである。

【0007】

【構成】本発明は、上記目的を達成するために、(1)対象言語の原文を入力するための入力部と、該入力部から入力された翻訳対象となる原文を記憶する原文記憶部と、翻訳処理に使用する知識情報を収容した翻訳辞書部と、該翻訳辞書部の内容を用いて前記入力された原文の翻訳処理を実行する翻訳部と、該翻訳部の翻訳処理の結果得られた訳文を記憶する訳文記憶部と、前記原文および訳文を表示する表示部と、これらを制御するとともに、前記表示部で表示された原文または訳文の編集処理を行う編集制御部とを備えた機械翻訳システムにおいて、前記翻訳部に自然な訳文を生成するために、原文である入力文の解析結果から得られる統語的情報と意味的情報に加えて変換ルールを用いて入力文を変換する変換手段を設け、該変換手段で変換された入力文を翻訳する

ようにしたこと、更には、(2) 前記変換ルールは、主語の意味素性から無生物主語に関するものを扱うようにしたことを特徴としたものである。以下、本発明の実施例に基づいて説明する。

【0008】図1は、本発明による機械翻訳システムの一例を説明するための構成図で、英日機械翻訳システムの全体ブロック図である。図中、1は形態素解析部、2は辞書、3は形態素形ファイル、4は第1の情報ファイル、5は構文解析部、6は文法規則、7は第1の解析ルール、8は英語構造同定部、9は第2の情報ファイル、10は第2の解析ルール、11は入力文変換ルール、12は第3の情報ファイル、13は変換部、14は第4の情報ファイル、15は構文生成部、16は形態素生成部である。本発明の実施例では、英語に多く見受けられる無生物主語の入力文を変換する場合を例にとって説明する。

【0009】機械翻訳システム中の辞書2には、語形、活用形、品詞、構造共起条件、訳語などが記述されている。図2は、動詞“allow”の辞書記述例を示す図である。図中、1行目は辞書の見出し語であり、2行目の+V(1)は“allow”が動詞であり、かつ無生物主語をとりやすいという情報を示すものである。無生物主語をとりやすい動詞には“+V(1)”というコードを、そうでない動詞には“+V(0)”というコードを付与しておく。無生物主語をとりやすい動詞としては、使役の意味を与える“cause”、“force”、“make”、“oblige”や、移動の意味を与える“bring”、“take”、“carry”、あるいは命令指示の意味を与える“tell”、“show”、“teach”といった動詞等がある。なお、ここでいう動詞とは一般動詞に限るもので、“be”動詞は除外して考える。また、図2において、2行目の残りの部分は活用形を示している。これより下の行に動詞“allow”の細分類がその構造共起条件とともに記載されており、各構造共起条件における訳語（および、その活用の模範）も併記されている。また、図3は名詞“version”の辞書記述例を示す図である。2行目にこの名詞が無生物のものであることを示す意味素性コード[-ANIMATE]が付与されている。

【0010】今、ここで“This version allows you to partition your hard disk.”なる英文が入力文として与えられたとする。入力された英文は、まず形態素解析部1で単語単位の解析が行われる。次に、構文解析部5によりどの単語とどの単語で句をなし、どの句がどの句、あるいは単語を修飾しているか等を知るために、文法規則（第1の解析ルール7）を順次適用して全ての可能な解析が取り出される。続いて、このような統語論的解析結果から、英語構造同定部8により第2の解析ルール10を用いて意味的におかしな解析が排除される。こうして、先の英文の解析結果として図4(a)に示すような木構造が得られる。以上のように、本発明の実施例

では英語構造同定部8中の一部として、入力英文の主語が持つ意味素性が[-ANIMATE]つまり無生物であれば、当該文中の主動詞とその細分類をキーとして入力文変換ルール11にアクセスして文変換を行うものである。

【0011】図5は、入力文変換処理のフローチャートである。以下、各ステップに従って順に説明する。

step1: まず、入力文の主語である名詞“version”の意味素性を見る。図3で見たように、この名詞は[-ANIMATE]という意味素性を持ち、無生物であることが判明する。主語の名詞が無生物でなければ、この処理は終了する。

step2: 主語が無生物であれば、次に、解析結果から入力文の主動詞の細分類を調べる。例文中の“allow”は、図2の「他動詞2」という細分類に当てはまることがわかる。

step3: 主動詞の見出し語と細分類コードにより変換ルールテーブルを検索する。すなわち、見出し語“allow”と細分類「他動詞2」という統語論的情報により、予め用意してある入力文変換ルールにアクセスする。図6に動詞“allow”の入力文変換ルールの記述例を示す。入力文変換ルールは動詞の見出し語とその細分類に応じて個々に用意してある。

step4: 所望の変換ルール②を適用して図4(b)のような木構造が得られ、この構造が変換部13へ送られる。

【0012】このように、解析結果が持つ統語論的かつ意味的情報に基づいて入力文変換ルールを適用して日本語構造への変換部に送られるので、訳出される日本語訳は、もはや無生物の主語による奇異な直訳調のものではなく、より自然なものになる。これは、英語の無生物主語は日本語では副詞的に訳されると日本語らしい表現になるという傾向を取り入れたものといえる。この結果、先の英文であれば、従来の訳文「このバージョンは、あなたがあなたのハードディスクを分割するのを許す」が「このバージョンにより、あなたはあなたのハードディスクを分割することができる」という訳文に改められることになる。

【0013】また、先に挙げた英文“Five minutes' walk will take me to the station.”の場合にも、主語の持つ意味素性が[-ANIMATE]つまり無生物であるので、主動詞“take”とその細分類をキーとして当該の入力文変換ルールにアクセスすれば“if I have five minutes' walk, I will get to the station.”なる英語文に変換される。この例文も実施例の英文と同様に、英語の無生物主語は日本語では副詞的に訳されると、日本語らしい表現になるという傾向を取り入れた変換例である。

【0014】

【効果】以上の説明から明らかなように、本発明によると、入力文の解析処理が終了した段階で、解析結果が持

つ統語的かつ意味的情報に基づき、変換規則を適用して入力文（原文）を変換し、変換された入力文を翻訳することにより、より自然な訳文を生成することができるので、訳文の理解容易性が向上し、ひいてはユーザが訳文をわかりやすくするために修正するといった負荷も大幅に軽減される。

【図面の簡単な説明】

【図1】 本発明による機械翻訳システムの一実施例を説明するための構成図である。

【図2】 動詞“allow”の辞書記述例を示す図である。

【図3】 名詞“version”の辞書記述例を示す図である。

【図4】 木構造を示す図である。

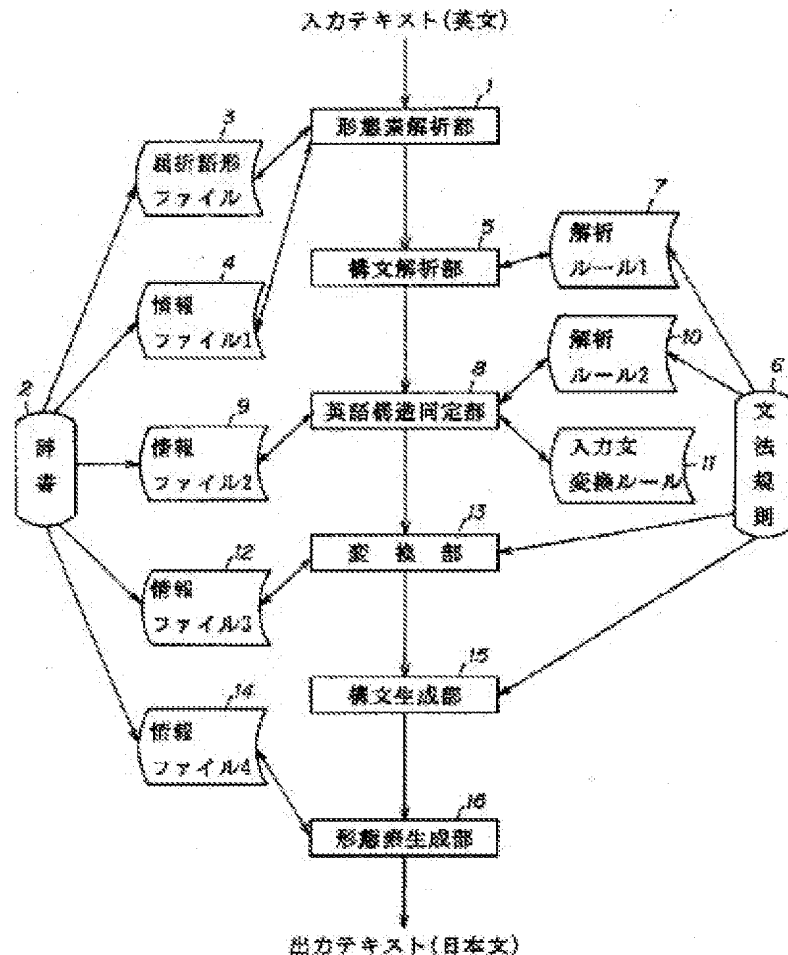
【図5】 入力文変換処理のフローチャートである。

【図6】 動詞“allow”の入力文変換ルールの記述例を示す図である。

【符号の説明】

1…形態素解析部、2…辞書、3…屈折語形ファイル、4…第1の情報ファイル、5…構文解析部、6…文法規則、7…第1の解析ルール、8…英語構造同定部、9…第2の情報ファイル、10…第2の解析ルール、11…入力文変換ルール、12…第3の情報ファイル、13…変換部、14…第4の情報ファイル、15…構文生成部、16…形態素生成部。

【図1】



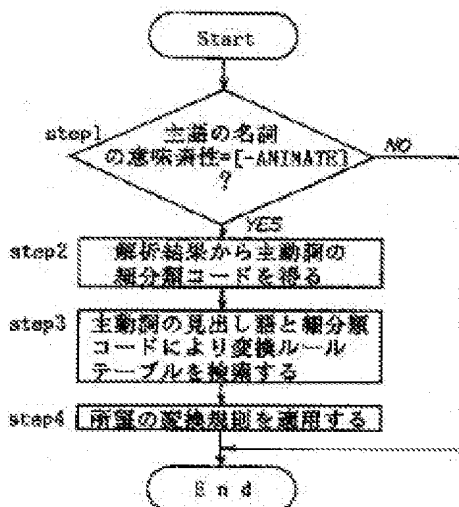
【図2】

allow		
+V(1), -ed -ed -s -ing		
自動詞		
/P(ol) + N		許す (動・サ5)
他動詞1		
/N		許す (動・サ5)
他動詞2		
/N + to-inf.		許す (動・サ5)
他動詞3		
/that-clause		認める (動・1)
.		.
.		.
.		.

【図3】

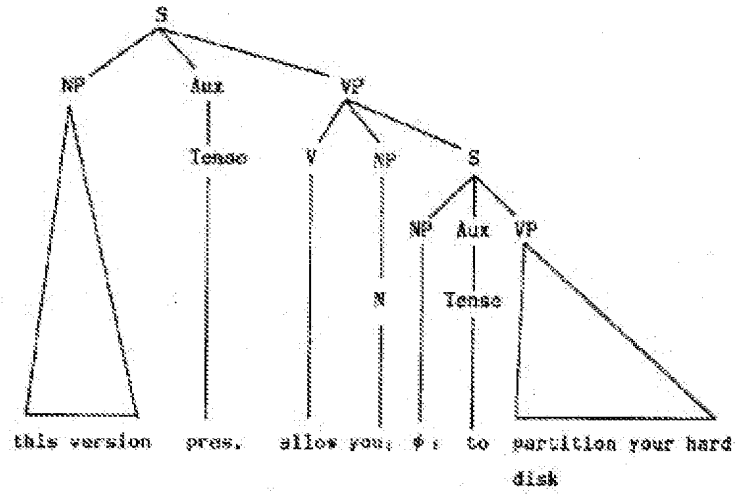
version	
+N[-ANIMATE], -s	
単数普通名詞	
バージョン (名)	
版 (名)	
翻訳 (名)	
説明 (名)	
脚色 (名)	
変種 (名)	
.	
.	
.	

【図5】

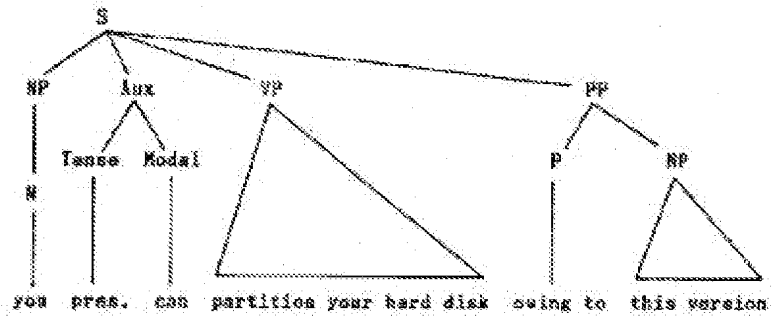


[圖4]

(a)



(b)



【図6】

見出し語	細分類コード	変換ルール記述
allow	他動詞 1'	① "allow"を"result in"に変換し、"allow"を支配している"V"ノードを"V"ノード("result"を直接支配)と"P"("in"を直接支配)ノードに枝分かれさせる。
	他動詞 2'	② "allow"と姉妹関係にある"NP"ノードが支配している終端記号列を1サイクル下の"S"ノードが直接支配している"NP"ノードの下にコピーする。次に、不定詞マーカの"to"を削除し、これを支配する"Aux"ノードを"Tense"ノードと"Modal"ノードに枝分かれさせ、"Tense"ノードの下には主動詞のテンス情報をコピーし、"Modal"ノードの下にはcan"を挿入する。次に、当該文の主語である"NP"ノードに前置詞"owing to"を付加して新しく"NP"ノードを作成して、1サイクル下の"S"ノードの最右端に直接付加する。この"S"ノードが支配する木構造のみを残すように枝刈りをする。